DOCUMENT RESUME

ED 185 896 HE 012 529

AUTHOR Sawyer, Richard; Maxey, E. James
TITLE The Validity Over Time of College Freshman Grade Prediction Equations.
INSTITUTION American Coll. Testing Program, Iowa City, Iowa. Research and Development Div.
REPORT NO ACT-RR-80
PUB DATE Oct 79
NOTE 19p.
AVAILABLE FROM ACT Publications, P.O. Box 168, Iowa City, IA 52243 ($1.00 prepaid)

EDRS PRICE MF01/PC01 Plus Postage.
DESCRIPTORS Academic Ability; *College Bound Students; *College Entrance Examinations; *College Freshmen; Females; Grade Point Average; *Grade Prediction; High School Graduates; Males; *Mathematical Formulas; Predictive Measurement; *Predictive Validity; Standardized Tests; Statistical Analysis
IDENTIFIERS *Act Assessment

ABSTRACT

The validity of prediction equations based on American College Testing (ACT) Program test scores and high school grades is investigated. Presented are separate results for prediction equations based on ACT scores only, on high school grades only, and on both kinds of prediction jointly. Also included is a discussion of the relationship between equations and the length of time they have been in use. The study is based on data collected from a random sample of 260 colleges participating in the ACT Research Services from 1972-73 to 1976-77. The accuracy of prediction equations based on the two variables was found to be quite stable over time. A similar stability was noted in the accuracy of grade predictions for males and females separately. Grade predictions based on ACT scores only showed slightly greater stability over time than those based on grades only. It is concluded that despite the changes some colleges experience in student abilities, curricula, and grading practices over time, freshman grade averages can be accurately predicted using equations as old as four years. ACT policy is to update equations at least every three years. Data tables and statistical analyses are briefly discussed, and a brief list of references is included. (Author/MSE)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

ED185896

THE MATERIAL IN ERIC DOCUMENTS IS SOMETIMES COPYRIGHTED MATERIAL. IT IS REPRODUCED IN ERIC BY PERMISSION OF THE COPYRIGHT HOLDER.

PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

U.S. DEPARTMENT OF HEALTH EDUCATION & WELFARE NATIONAL INSTITUTE OF EDUCATION

Prepared by the Research and Development Division
The American College Testing Program

For additional copies write:
ACT Publications
P.O. Box 168
Iowa City, Iowa 52243

(Check or money order must accompany request.) Price: $1.00

© 1979 by The American College Testing Program. All rights reserved.

ABSTRACT

This report investigates the validity of prediction equations based on ACT test scores and high school grades. It presents separate results for prediction equations based on ACT test scores only, on high school grades only, and on both kinds of prediction jointly. It also includes a discussion of the relationship between the validity of the prediction equations and the length of time they have been in use.

The study is based on data collected from a random sample of 260 colleges that participated in the ACT Research Services from 1972-73 to 1976-77. Separate prediction equations for each college were calculated from data for the years 1972-73, 1973-74, 1974-75, and 1975-76. These prediction equations were then applied to data from 1976-77 freshmen, and the predicted and actual grades were compared.

The accuracy of the prediction equations based on ACT test scores and high school grades jointly was quite stable over time. The mean absolute error of prediction ranged from .53 to .55 over the four years; the proportion of students whose predicted grade was within .20 of their earned grade ranged from .24 to .25; and the cross-validated correlation ranged from .55 to .56. A similar stability was noted in the accuracy of grade predictions for males and females separately.

Grade predictions based only on ACT test scores had a slightly smaller mean absolute error than grade predictions based only on high school grades, and showed slightly greater stability over time. The mean absolute error of predictions based only on test scores was about .57 for the whole four-year period. For predictions based only on high school grades, the mean absolute error ranged from .57 to .60 over the four-year period.

One can conclude that although most colleges experience some change in their students' academic abilities, their curricula, or their grading practices over a period of time, freshman grade average can in most cases be accurately predicted using equations as old as four years. It is ACT policy that colleges update their prediction equations at least every three years if they are to receive grade predictions for future applicants. The data from this study suggest that more frequent revision is not generally necessary. In considering whether revision of equations is necessary before the required date, college researchers should, of course, determine whether some change has occurred which would necessitate an early revision.

THE VALIDITY OVER TIME OF COLLEGE FRESHMAN GRADE PREDICTION EQUATIONS

Richard Sawyer
E. James Maxey

A problem commonly encountered in predicting college freshman grades from standardized test scores and high school grades is the validity over time of the prediction equations. Changes over time in the distribution of ability among entering students, in a college's entrance requirements, in its freshman curriculum, and in instructors' grading policies can be great enough to make old prediction equations inaccurate. Deterioration of the accuracy of grade predictions has obvious negative implications for both colleges and students. On the other hand, collecting and reporting the data needed to revise prediction equations can be expensive and time-consuming. Therefore, many college researchers would naturally want to use prediction equations as long as possible before updating them.

To accommodate the time schedules of colleges, ACT updates its prediction equations each fall using data collected from students who were freshmen in the previous year. Prediction equations are, therefore, based on freshman grades that are at least one year old. Because most students who take the ACT Assessment in the fall of one year will be college freshmen in the fall of the next year, prediction weights are typically calculated from college grades that are at least two years older than the grades being predicted. To minimize the error from out-of-date equations, ACT requires colleges to participate in its predictive research services at least once every three years if they wish to continue to receive grade predictions for future applicants. Thus, predicted grades are typically based on data that are two to four years older than the grades being predicted.

The primary purpose of this report is to document the relationship between the validity and the age of grade prediction equations based on ACT Assessment test scores and high school grades. A second purpose is to summarize the validity of locally developed grade prediction equations based on ACT test scores only, on high school grades only, and on both kinds of predictors jointly. A third purpose is to suggest various techniques for evaluating local prediction equations.

Prediction equations were calculated for a probability sample of individual colleges using data from the years 1972-73, 1973-74, 1974-75, and 1975-76. The predicted grades from these equations were then compared with the actual freshman grades earned in 1976-77, and the comparisons are stated in terms that are easily interpreted and used.

Previous Research

There is relatively little published research on the rate at which predictive accuracy declines over time. Hills, Klock, and Bush (1965) compared the predicted and earned grades of students at seven colleges in Georgia over a three-year period. They found that the average correlation between the predicted freshman grade average and the actual grade average one and two years later was .64 and .63, respectively. The average correlation between freshman grade average and predictors in the base year was .67.

Bowers and Loeb (1972) found that in predicting grades for freshmen at the University of Illinois, the weights for ACT Composite scores were unstable over a five-year period. They found that the weight for high school percentile rank as a predictor of freshman grades was more stable.

Perrin and Whitney (1976) studied the ACT scores, high school grades, and freshman grades from a national sample of student records. They found very little difference in the accuracy of expectancy tables two and three years older than the freshman grade expectancies they were predicting.

The ACT Assessment Program

The ACT Assessment Program is a comprehensive evaluative, guidance, and placement service for students and educators involved in the transition of students from high school to college. This program is based on the ACT Assessment, which consists of four academic tests, self-reported high school grades, the Student Profile Section (SPS), and the ACT Interest Inventory. The program also includes a number of supporting research services, described below.

The four academic tests of the ACT Assessment measure abilities in the subject areas traditionally identified with college and high school programs: English, mathematics, social studies, and natural sciences. The English Usage Test measures students' understanding and use of the basic elements of correct and effective writing; the Mathematics Usage Test measures their mathematical reasoning and problem-solving ability; the Social Studies Reading Test measures the problem-solving skills required in the social studies; the Natural Science Reading Test measures the critical reasoning and problem-solving skills required in the natural sciences. The arithmetic average of the scores on these four tests is the ACT Composite score, which is often used as a measure of overall academic ability. ACT test scores are reported on a standard scale that ranges from 1 to 36. More detailed descriptive and technical information about ACT test scores can be found in the *Technical Report for the ACT Assessment Program* (1973).

When students register for the ACT Assessment Program, they report the last grade received in each of the four subject areas prior to the senior year of high school. The arithmetic average of these four grades—defined as the high school average—provides another measure of overall academic ability. Maxey and Ormsby (1971) investigated the accuracy of self reported high school grades and found that about 78% of the students reported their grades correctly.

Another component of the ACT Assessment Program is the Student Profile Section (SPS). Through the SPS, students provide information about their background, extracurricular accomplishments, special academic needs, housing plans, financial need, planned major, and career plans. Students complete the SPS when they register for the ACT Assessment.

The ACT Interest Inventory measures students' preferences for job-related activities in six basic interest dimensions: Science, Creative Arts, Social Service, Business Contact, Business Detail, and Technical. Students complete the Interest Inventory when they register for the ACT Assessment.

Information from all these sources—the ACT tests, high school grades, SPS, and Interest Inventory—is organized into individual Student Profile Reports sent to students and colleges. With the summary information provided by the Student Profile Report, students and educators can make informed decisions and plans.

The ACT Predictive Research Services

ACT offers without charge to colleges two general plans for predicting freshman grades. Each plan is designed to meet the varying needs and resources of colleges which use ACT data.

The Basic Research Service requires minimal effort by a college in reporting data. To participate in the Basic Research Service, college personnel simply mark on a computer-generated roster the overall grade point averages of each of their freshmen. A minimum data base of 100 records is required. Through the Basic Research Service, a college can obtain predictions of overall freshman grade average based on the four ACT test scores and the high school average. If they have data for at least 100 students of each sex, colleges can also obtain separate prediction equations for males and females. During the 1977-78 academic year, 461 colleges reported grades for 126,880 freshmen through the Basic Research Service.

The Standard Research Service is designed for colleges that want to predict specific course grades, develop prediction equations for sub-

groups of students, or use predictors other than ACT test scores and high school grades. A large variety of research studies can be accommodated by the Standard Research Service. Data can be supplied either through punched cards, optically scanned cards, or magnetic tape. During 1977-78, 185 colleges reported grades for 150,998 students through the Standard Research Service.

Data Base

This study is based on a sample from a data base consisting of student records submitted by institutions through their participation in ACT's predictive research services. The institutions represented in this data base participated in the ACT predictive research services in the academic year 1976-77 and in one or more of the academic years 1972-73, 1973-74, 1974-75, and 1975-76. Therefore, the grades earned by the 1976-77 freshmen at these institutions were available for comparison with grades predicted from equations developed in one or more of the four preceding years. There are 605 colleges represented in the data base.

Because the data in the study were collected from colleges participating in ACT's predictive research services, in some respects they are not representative of students nationally:

- Colleges using the ACT Assessment are located mainly in the Rocky Mountains, Great Plains, South, and Midwest with comparatively fewer in the East/Northeast and on the West Coast.
- Privately controlled institutions are relatively underrepresented among colleges that use the ACT Assessment, and publicly controlled institutions are overrepresented.
- Because participation in ACT's research services is voluntary, the data base is self-selected even among colleges that use the ACT Assessment Program.

The results of the study are, therefore, not necessarily representative of the results that would be obtained if data from all colleges in the nation could somehow be collected. One should be cautious, therefore, in applying the results to institutions which do not use the ACT Assessment or do not participate in ACT's predictive research services. Nevertheless, the study will suggest major trends and extend knowledge in this area beyond the results available to date.

To maximize the number of colleges from which data were available, the variables used in the predictions were restricted to the following specifications: overall grade average as criterion, high school grades and ACT scores as predictors, and subgrouping on sex.

Most of the grade averages in this study are from the first semester of the freshman year. Colleges participating in some of ACT's research services do have the option of pooling grades from previous years or reporting grade averages based on the entire freshman year. ACT does not maintain records of individual colleges' choice of criteria. However, from examining production volumes throughout the year, we estimate that over 60% of the colleges in the data base reported first semester grades for the current academic year, and the rest either reported first year cumulative grade averages or pooled data from previous years.

Sample Design

To reduce the computational costs of this study, weights were calculated and prediction equations were cross-validated on a probability sample of records selected from the above data base. The sampling was carried out in two stages.

First, a random sample of 260 colleges was selected from the 605 colleges in the data base. Weights were computed from all student records submitted by these 260 institutions in the academic years 1972-73, 1973-74, 1974-75, and 1975-76. Observe that not every institution supplied data in every year. The number of colleges by year in the data base and sample is displayed in Table 1.

TABLE 1

Summary of Data Base and Sample for Cross-Validation Study

Year	Number of colleges in data base	Number of colleges in sample	Number of student records for computing weights	Number of 1976-77 student records for cross-validation
1972-73	451	187	97,985	13,732
1973-74	484	203	114,331	15,589
1974-75	494	207	108,118	15,412
1975-76	520	211	109,207	15,384
1976-77	605	260	—	—

For each college in the sample, a random subsample of the 1976-77 student records was then selected. The differences between actual 1976-77 grades and predictions based on 1972-73 equations, 1973-74 equations, and so on, were summarized. Thus a cross-validation was made for all of the equations developed from data in the four years preceding 1976-77.

At the time the study was done, 1976-77 grades were the most recent available for cross-validating prediction equations developed from earlier years' data. Computational costs prevented replicating the study using earlier base years and cross-validation years.

It was anticipated that the validity of grade predictions might vary according to differences in the composition and affiliation of the colleges. By taking note of these differences when designing the sample, it was hoped that greater precision in the inferences might be obtained. In this study, separate samples were selected from strata of colleges determined by their control, type, and size. The stratification variables were:

- The *control* of a college, public or private.
- The *type* of a college, as determined by the highest degree level it offers:

  Type 1: Two but less than four years of work beyond Grade 12—includes junior colleges, technical institutes, normal schools.

  Type 2: Only the bachelor's or first professional degree—includes those institutions offering courses of study leading to the customary Bachelor of Arts or Bachelor of Science degree and all those degrees which entitle the possessor to enter the profession indicated.

  Type 3: Master's and/or second professional degree—includes those institutions offering the customary first graduate degree and any degree earned in the same field after the bachelor's or first professional degree. This type of institution does *not* offer the Doctor of Philosophy or equivalent degrees.

  Type 4: Doctor of Philosophy and equivalent degrees—includes those institutions which are considered universities.

- The *size* stratum for a college, as determined by the number of students for which the college reported 1976-77 freshman grades:

Category 1: 100 students or fewer

Category 2: 101-200 students

Category 3: 201-500 students

Category 4: 501-1000 students

Category 5: 1001+ students

Colleges do have the option of reporting grades for a sample of their freshman class, provided the sample size is 100 or larger. Moreover, colleges can pool data from previous years in order to attain the required minimum sample size. Therefore, the size measure in this study is not necessarily the same as total freshman class size. However, for about 70% of the institutions, the number of records in the data base is within 30% of the estimated freshman enrollment.

The subsampling rates for 1976-77 records were a function of the size strata for the college sample. These subsampling rates and the number of colleges selected from each size stratum are displayed in Table 2.

These sample sizes were chosen to yield a 95% chance that the mean absolute error (defined below) estimated from the sample would be within the following limits of the mean absolute error computed from all records in the data base:

- ± .01 grade units of the mean absolute error of all students in all colleges.
- ± .03 grade units of the mean absolute error for all students in each size stratum.

The second condition was imposed to permit the use of the sample to study the effects of college size on predictive accuracy (results not reported here). Examination of the sampling variances estimated from the data indicated that these precisions were attained.

TABLE 2

Stratification and Subsampling for the Cross-Validation Study Sample by Size Stratum

Size stratum	Number of colleges in data base	Number of colleges in sample	Subsampling rate for 1976-77 records
100 or fewer	47	38	1/2
101-200	221	75	1/4
201-500	202	63	1/4
501-1000	70	44	1/6
Over 1000	65	40	1/16
Total	605	260	—

Prediction Equations

Prediction equations were calculated from the 1972-73, 1973-74, 1974-75, and 1975-76 data using a standard four-variable multiple linear regression of college freshman grade average on the four ACT test scores:

$$y = a_0 + a_1 \cdot \text{ACT English score} + a_2 \cdot \text{ACT Mathematics score} + a_3 \cdot \text{ACT Social Studies score} + a_4 \cdot \text{ACT Natural Science score}$$

where a_0, a_1, a_2, a_3, and a_4 are regression weights calculated from the base year data. In the ACT Standard Research Service this prediction is called the T Index.

Prediction equations were also calculated using:

- the four self-reported high school grades only in a multiple linear regression equation (H Index).
- the average of the T Index and H Index (the TH Index).
- the four ACT test scores and the four high school grades in an eight-variable multiple linear regression equation.

Past research has shown that the correlation between the TH Index and freshman college grade average is only slightly less than that for the traditional eight-variable multiple linear regression equation (*Technical Report for the ACT Assessment Program*, 1973).

Previous research also suggests that the relationship between high school grades, standardized test scores, and freshman grade average differs for males and females (*Technical Report for the ACT Assessment Program*, 1973). Moreover, colleges often elect to receive separate predictions by sex. For these reasons, prediction equations were calculated separately for males and females in each college, as well as for all students in a college.

Some colleges in the sample submitted student grade data without identifying the sex of their students. Therefore, the sum of the number of males and females from which the separate-sex equations were computed is less than the total number of records from which the combined-sex equations were computed.

Cross-Validation Statistics

The difference between the predicted and actual 1976-77 grade average for a student is called the error of prediction. The magnitude of this error, ignoring the direction, is called the absolute error of prediction. If the absolute error of prediction is averaged over a group of students (for example, all the students in a college), the resulting number is called the mean absolute error (MAE) of prediction in that group of students. This statistic has an immediate relevance to the quality of grade predictions. For example, if the MAE in predicting grade average in a college is .45, then, on the average, there is an absolute discrepancy of .45 grade units between the predicted and actual grade averages in the college.

A related measure of prediction quality that is sometimes used is the mean squared error, which is the average over a group of students of the squared error of prediction. It is actually this quantity that the usual least-squares regression methods seek to minimize. Because the mean squared error is less directly interpretable in practical terms than the MAE, it was not used here.

Another useful measure of the quality of prediction for a group of students is the proportion whose predicted grade is within a certain range of the actual grade. In this study we report the proportion of students whose predicted grade average is within 0.20 grade units of actual grade average; this proportion will be denoted by the abbreviation P20. An absolute error of 0.20 grade units represents a high degree of accuracy in prediction. The statistic P20, therefore, measures the proportion of students for whom extremely accurate grade predictions were possible. Some researchers may consider prediction errors larger than 0.20 to be quite

acceptable; analogous proportions (such as the proportion of students whose predicted grade average is within 0.50 grade units of the actual grade point average) could be computed for them.

Probably the most commonly used measure of predictive efficiency is "cross-validated r," that is, the Pearson correlation between the predicted and actual grades. Generally, this coefficient is compared with the correlation coefficient calculated from the base year data to give an indication of the accuracy and stability of the prediction equations. Because it is so widely used, this statistic is reported in addition to MAE and P20.

A limitation of this statistic is illustrated by the hypothetical data in Figures 1a and 1b. These plots suggest that false conclusions could result from using cross-validated r as the sole measure of prediction quality. In both cases, cross-validated r is near 1, yet in both cases the prediction is not very good. In the first instance, the predicted grades are consistently too low; in the second, they are systematically too low for low-ranking students and too high for high-ranking students. Thus, a large cross-validated r does not guarantee success in prediction. A low value of cross-validated r will indicate poor prediction, however, because if the predicted and actual grades have little linear relationship, they cannot be very close to each other.

A justification sometimes advanced for the correlation coefficient is that r^2 measures the proportional decrease in squared error from that which would result if the mean grade were predicted for each student. Similarly, r^2 takes into account the effect on prediction due to the inherent variability of students' grades.

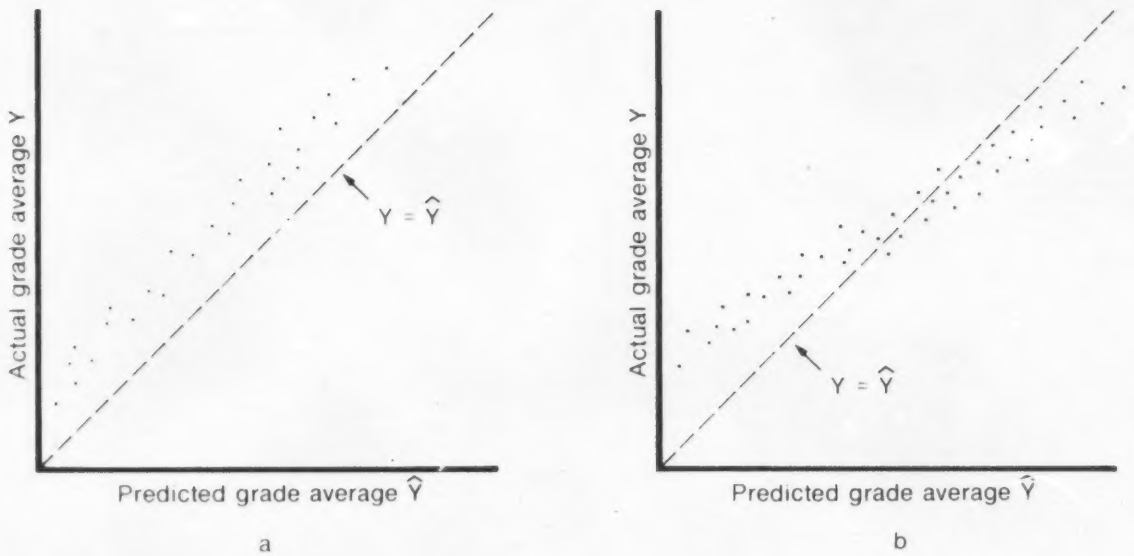

Figure 1. Two situations in which cross-validated r gives a false indication of prediction quality

Computations

The prediction equations and cross-validation statistics for each college were computed by a custom program using the IMSL subroutine package (IMSL, 1977). The program was tested on data from three colleges over a three-year period. Every computation on these data was compared with one obtained from the SPSS statistical package (Nie, Hall, Jenkins, Steinbrenner, and Brent, 1975) and, where feasible, with one computed by hand. In all cases, the cross-validation statistics for each college were within ± .001 of each other, and in a majority of cases they were within ± .0001.

The cross-validation statistics from each college were summarized through SPSS routines. The data were weighted to take into account the differential sampling rates in the various strata.

Results

The results of the analyses are displayed in Tables 3a-3d, 4a-4d, 5, and 6. They indicate that for *students* as a whole (irrespective of their college), the accuracy of all four kinds of prediction equations was quite stable over the four-year period studied. The same is generally true of the accuracy of the prediction equations for *colleges*, although a few individual colleges do show some instability.

Tables 3a-3d display the cross-validation statistics for all students in the sample. Over the four year period, there is no marked increase or decrease in any of the three statistics for the four kinds of predictions. For the T Index, MAE for the total group of students is about .57, P20 is .22-.23, and r is .48-.49. For the H Index, there is slightly more deterioration over time: MAE ranges from .57-.60, P20 from .21-.23, and r from .49-.51. The TH Index predictions have an MAE of .54-.55, a P20 of .23-.24, and an r of .55-.56. The eight-variable regression equations showed virtually the same results: MAE from .53-.55, P20 from .24-.25, and r from .55-.56.

TABLE 3a

Accuracy of Prediction of 1976-77 Student Grade Average from Equations Developed in Years 1972-73 through 1975-76

(T Index—ACT Test Scores Only)

Group	Cross-validation statistic	Year in which prediction equation was developed			
		1972-73	1973-74	1974-75	1975-76
Total group	MAE	.57	.57	.57	.57
	P20	.23	.22	.23	.23
	r	.48	.48	.49	.49
Males	MAE	.61	.61	.60	.61
	P20	.21	.20	.21	.21
	r	.43	.43	.43	.43
Females	MAE	.56	.55	.55	.55
	P20	.22	.24	.25	.23
	r	.53	.50	.51	.51

TABLE 3b

Accuracy of Prediction of 1976-77 Student Grade Average from Equations Developed in Years 1972-73 through 1975-76

(H Index—High School Grades Only)

Group	Cross-validation statistic	Year in which prediction equation was developed			
		1972-73	1973-74	1974-75	1975-76
Total group	MAE	.60	.59	.57	.57
	P20	.21	.23	.22	.23
	r	.49	.49	.50	.51
Males	MAE	.62	.61	.60	.60
	P20	.21	.20	.22	.21
	r	.46	.46	.46	.48
Females	MAE	.57	.55	.54	.54
	P20	.24	.23	.24	.24
	r	.50	.53	.54	.53

TABLE 3c

Accuracy of Prediction of 1976-77 Student Grade Average from Equations Developed in Years 1972-73 through 1975-76

(TH Index)

Group	Cross-validation statistic	Year in which prediction equation was developed			
		1972-73	1973-74	1974-75	1975-76
Total group	MAE	.55	.55	.54	.54
	P20	.23	.23	.24	.24
	r	.55	.55	.56	.56
Males	MAE	.59	.59	.57	.58
	P20	.22	.22	.23	.23
	r	.50	.50	.50	.51
Females	MAE	.53	.52	.52	.52
	P20	.24	.24	.26	.25
	r	.58	.58	.59	.59

TABLE 3d

Accuracy of Prediction of 1976-77 Student Grade Average from Equations Developed in Years 1972-73 through 1975-76

(Eight-Variable Multiple Linear Regression)

Group	Cross-validation statistic	Year in which prediction equation was developed			
		1972-73	1973-74	1974-75	1975-76
Total group	MAE	.55	.54	.53	.54
	P20	.24	.24	.25	.24
	r	.55	.55	.56	.56
Males	MAE	.59	.59	.58	.58
	P20	.22	.23	.23	.23
	r	.50	.50	.50	.51
Females	MAE	.53	.52	.51	.51
	P20	.24	.25	.26	.25
	r	.57	.57	.59	.59

The T Index predictions were somewhat more stable over time than the H Index predictions. The reasons for this cannot be easily determined, but could be related to changes in high school curriculum or inflation in high school grades.

Although the H Index r's are slightly but constantly larger than the T Index r's, the statistics MAE and P20 indicate that, on the average, the T Index resulted in slightly better predictions than did the H Index. That the H Index r is larger than the T Index r is consistent with earlier published research (*ACT Technical Report*, 1973). The results for MAE and P20, however, suggest that something like the situation illustrated in Figures 1a and 1b occurred in many of the colleges.

Further examination of the prediction errors revealed that the H Index predictions based on the 1972-73, 1973-74, 1974-75, and 1975-76 data underestimated the criterion on the average by .14, .10, .04, and less than .01, respectively. The T Index predictions based on data from these four years often overestimated the criterion, but by an average of only .05, .05, .05, and .03, respectively. Thus, the H Index tended to underestimate the criterion, and the amount of underestimation increased with the age of the prediction equation. The T Index tended to overestimate the criterion, but by a fairly constant and small amount over time.

Prediction equations developed for males and females separately show the same stability in quality as the total group equations. The predictions for females were somewhat more accurate (eight-variable multiple regression MAE = .51-.53, P20 = .24-.26, r = .57-.59) than the predictions for males (eight-variable multiple regression MAE = .58-.59, P20 = .22-.23, r = .50-.51). This result is consistent with earlier analyses of ACT grade predictions (*ACT Technical Report*, 1973).

It should be noted that these results pertain to separate-sex prediction equations, rather than to a total group equation. The cross-validation results for predictions made from combined-sex equations, however, are virtually identical to those for predictions made from the separate-sex equations. This would indicate that there is little average benefit in developing separate-sex equations.

Tables 4a-4d display the college medians of the cross-validation statistics. The numbers in this table show that measures of prediction quality for students across all colleges are also typical of the measures of prediction quality for individual colleges. The same stability in prediction quality over the four years is evident.

TABLE 4a

Median College Cross-Validation Statistics for Prediction Equations Developed from Data in Years 1972-73 through 1975-76

(T Index—ACT Test Scores Only)

Group	Cross-validation statistic	Year in which prediction equation was developed			
		1972-73	1973-74	1974-75	1975-76
Total group	MAE	.55	.56	.55	.55
	P20	.23	.23	.22	.22
	r	.48	.48	.48	.48
Males	MAE	.60	.61	.60	.60
	P20	.22	.24	.20	.20
	r	.44	.42	.44	.42
Females	MAE	.54	.54	.53	.54
	P20	.23	.24	.24	.23
	r	.50	.52	.53	.51

TABLE 4b

Median College Cross-Validation Statistics for Prediction Equations Developed from Data in Years 1972-73 through 1975-76

(H Index—High School Grades Only)

Group	Cross-validation statistic	Year in which prediction equation was developed			
		1972-73	1973-74	1974-75	1975-76
Total group	MAE	.58	.57	.60	.54
	P20	.21	.21	.23	.23
	r	.48	.50	.49	.50
Males	MAE	.60	.62	.59	.59
	P20	.20	.20	.21	.22
	r	.45	.47	.45	.48
Females	MAE	.57	.54	.53	.53
	P20	.21	.22	.24	.23
	r	.54	.53	.54	.55

TABLE 4c

Median College Cross-Validation Statistics for Prediction Equations Developed from Data in Years 1972-73 through 1975-76

(TH Index)

Group	Cross-validation statistic	Year in which prediction equation was developed			
		1972-73	1973-74	1974-75	1975-76
Total group	MAE	.53	.54	.52	.52
	P20	.24	.23	.24	.24
	r	.55	.56	.56	.56
Males	MAE	.56	.59	.56	.56
	P20	.22	.21	.23	.23
	r	.54	.54	.53	.55
Females	MAE	.52	.51	.50	.50
	P20	.24	.24	.25	.24
	r	.58	.60	.60	.59

TABLE 4d

Median College Cross-Validation Statistics for Prediction Equations Developed from Data in Years 1972-73 through 1975-76

(Eight-Variable Multiple Linear Regression)

Group	Cross-validation statistic	Year in which prediction equation was developed			
		1972-73	1973-74	1974-75	1975-76
Total group	MAE	.53	.53	.52	.51
	P20	.24	.24	.25	.25
	r	.54	.55	.56	.56
Males	MAE	.57	.60	.57	.56
	P20	.21	.21	.21	.23
	r	.52	.51	.51	.51
Females	MAE	.52	.51	.50	.52
	P20	.25	.25	.26	.25
	r	.57	.57	.5	.58

Table 5 presents a frequency distribution of MAE for the eight-variable regression among colleges for the four years studied. For about 72%-78% of the colleges, the MAE is 0.60 or less. Thus, the MAE reported in Table 3d for students over all colleges is fairly typical of the MAEs in most colleges individually. For a small proportion (2%-4%) of colleges, MAEs in the range 0.80-1.10 occurred.

The relative stability of the quality of predictions over all students and colleges could mask instability in particular colleges. Table 6, which addresses this issue, is a frequency distribution of the differences between MAE for eight-variable multiple regression equations developed in 1975-76 and MAE for the equations developed in 1974-75, 1973-74, and 1972-73. The results indicate that for most colleges, MAE is quite stable from year to year. The number of colleges with a larger MAE from 1974-75 prediction equations than from 1975-76 prediction equations is approximately the same as the number of colleges with a smaller MAE from 1974-75 prediction equations than from 1975-76 equations. Moreover, about 99% of the differences in MAE are in the range -.10 to +.10. Slightly more colleges have a larger MAE from equations developed in 1974-75, 1973-74, or 1972-73 than 1975-76, but the differences in MAE are concentrated in the range of -.10 to .00. In summary, the MAE for over 90% of the colleges remains stable, even over a period of four years.

TABLE 5

Proportion of Colleges with Various Ranges of Mean Absolute Error in Predicting 1976-77 Freshman Grade Point Average

(Eight-Variable Multiple Linear Regression)

Range in MAE	Year in which prediction equation was developed			
	1972-73	1973-74	1974-75	1975-76
0.0 - 0.1	.00	.00	.00	.00
0.1 - 0.2	.00	.00	.00	.00
0.2 - 0.3	.00	.00	.00	.00
0.3 - 0.4	.07	.08	.12	.09
0.4 - 0.5	.31	.31	.29	.36
0.5 - 0.6	.34	.34	.37	.33
0.6 - 0.7	.18	.21	.15	.13
0.7 - 0.8	.09	.05	.06	.05
0.8 - 0.9	.02	.02	.01	.02
0.9 - 1.0	.00	.00	.00	.01
1.0 - 1.1	.00	.00	.00	.01

TABLE 6

Stability of Mean Absolute Error across Four Years: Proportions of Colleges with Various Differences in MAE from 1975-76 Equations and MAE from Equations Developed in Earlier Years

(Eight-Variable Multiple Linear Regression)

Range of difference in MAE from 1975-76 equations and MAE from older equation	Earlier year in which prediction was developed		
	1972-73	1973-74	1974-75
-0.50 to -0.40	.00	.00	.00
-0.40 to -0.30	.00	.00	.00
-0.30 to -0.20	.00	.00	.00
-0.20 to -0.10	.06	.03	.01
-0.10 to 0.00	.63	.69	.55
0.00 to 0.10	.30	.28	.44
0.10 to 0.20	.00	.00	.00
0.20 to 0.30	.01	.00	.01
0.30 to 0.40	.00	.01	.00
0.40 to 0.50	.00	.00	.00
0.50 to 0.60	.00	.00	.00

Summary and Conclusions

The stability of prediction equations based on ACT scores and high school grades was investigated for a stratified random sample of 260 colleges that participated in the ACT Research Services during the period of 1972-73 to 1976-77. Separate prediction equations for each college were calculated from data for the years 1972-73, 1973-74, 1974-75, and 1975-76 and applied to 1976-77 freshmen. Prediction errors for 1976-77 grade averages were then summarized.

The results indicate that for most students and colleges, the quality of the prediction equations is quite stable over the four years. The mean absolute error of prediction using an eight-variable multiple regression equation ranged from .53 to .55; the proportion of students whose predicted grade was within .20 of their earned grade ranged from .24 to .25; and the cross-validated correlation ranged from .55 to .56. A similar stability was noted in the accuracy of predictions for males and females using separate-sex prediction equations.

Grade predictions based on test scores only had a slightly smaller mean absolute error than grade predictions based on high school grades only and showed slightly greater stability over time. The mean absolute error of predictions based on test scores only was about .57 for the whole four year period. For predictions based on high school grades only, it ranged from .57 to .60 over the four year period.

One can conclude from this study that although most colleges experience some change in their students' academic abilities, their curricula, or grading practices over a period of time, freshman grade average can in most cases be accurately predicted using equations which have been in use for as long as four years. It is ACT's policy that colleges update their prediction equations at least every three years if they are to receive grade predictions for future applicants. The data from this study suggest that more frequent revision is not generally necessary. In considering whether revision of equations is necessary before the required date, college researchers should, of course, determine whether some change has occurred which would necessitate an early revision.

REFERENCES

The American College Testing Program. *Technical report for the ACT Assessment Program.* Iowa City, Iowa: Author, 1973.

Bowers, J., & Loeb, J. Changes in predictor weights in GPA regression equations. *Measurement and Evaluation in Guidance,* 1972, *5,* 438-442.

Hills, J., Klock, J., & Bush, M. The use of academic prediction equations with subsequent classes. *American Educational Research Journal,* 1965, *2,* 203-206.

International Mathematical and Statistical Laboratory. *IMSL library 1 reference manual* (6th ed.). Houston, Texas: Author, 1977.

Maxey, E. J., & Ormsby, V. J. *The accuracy of self-report information collected on the ACT test battery: High school grades and items of nonacademic achievement* (ACT Research Report No. 45). Iowa City, Iowa: The American College Testing Program, 1971.

Nie, N., Hull, C. H., Jenkins, J., Steinbrenner, K., & Bent, D. *Statistical package for the social sciences* (2nd ed.). New York: McGraw-Hill, 1975.

Perrin, D. W., & Whitney, D. R. Methods for smoothing expectancy tables applied to the prediction of success in college. *Journal of Educational Measurement,* 1976, *13,* 223-231.